DOCUMENT RESUME

ED 035 271 EM 007 594

AUTHOR Coulson, J. E.; Silberman, H. F.
TITLE - Results of Initial Experiment in Automated Teaching.
INSTITUTION System Development Corp., Santa Monica, Calif.
REPORT NO SP-73
PUB DATE 8 Jul 59
NOTE 33p.
AVAILABLE FROM System Development Corporation, 2500 Colorado Ave., Santa Monica, California 90404

EDRS PRICE EDRS Price MF-$0.25 HC Not Available from EDRS.
DESCRIPTORS *Autoinstructional Aids, Constructed Response, Educational Research, Educational Technology, Instructional Technology, Linear Programing, *Optional Branching, Pacing, *Programed Instruction, Programed Materials, Program Evaluation, Program Improvement, Programing, *Response Mode, Sequential Programs, *Teaching Machines

ABSTRACT

Eight groups of ten junior college students were given an experimental training session with manually simulated teaching machines, each group being taught with a different mode of teaching machine operation. The variables were student response mode, size of steps between successive items, and branching procedure. A written test was given to all students after the training session, and again three weeks later. The same test was given to a control group which had no training with the teaching machines. The use of the simulated teaching machines led to a significant degree of learning by the subjects. Multiple choice and branching procedures required less time but did not affect test scores significantly. Small item steps required significantly more training time, but also yielded significantly higher test scores. Significant interaction between the mode of response and branching variables resulted from a high mean criterion score obtained by the constructed response--non-branching group. Statistical tables supplement the report. [Not available in hard copy due to marginal legibility of original document.] (JY)

INSTRUCTIONAL TECHNOLOGY - U.S.C.

ED035271

PROCESS WITH MICROFICHE AND PUBLISHER'S PRICES. MICROFICHE REPRODUCTION ONLY.

SP - 73

RESULTS OF INITIAL EXPERIMENT IN AUTOMATED TEACHING

J. E. Coulson and H. F. Silberman

8 July 1959

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA, CALIF.

EM007 594

PROCESS WITH MICROFICHE AND PUBLISHER'S PRICES. MICRO-FICHE REPRODUCTION ONLY.

ACKNOWLEDGEMENTS

The authors of this document wish to acknowledge the assistance of the following people in the numerical analysis of the data, for their helpful suggestions, or for time spent working on the project at various intervals: C. M. Bertone, S. Cooney, R. E. Dear, W.R.Ellis,Jr., K. Esgate (Santa Monica City College), C. Fuller, T.E. Henderson, B.T. Jensen, S.L. O'Connell, B. Patton, C. D. Parkinson, M. S. Rogers, Miss Matilda Sims, L. Staugas, W. Thomas (Santa Monica City College), R. J. Tuber.

RESULTS OF INITIAL EXPERIMENT IN AUTOMATED TEACHING

Interest in automated teaching has increased greatly in the past ten years, as evidenced by the number of papers on the subject given in recent psychological and educational conferences and by the variety of teaching machines now available for research or commercial purposes. This new interest has been stimulated by the critical shortage of skilled teachers throughout the country, and by the belief of many psychologists and educators that automated teaching devices can be a valuable tool for teachers in both academic and industrial situations. For some learning situations the teaching machine may surpass its human counterpart, particularly where precise control of the timing and contingencies for student reinforcement is critical.

A "teaching machine" may be defined generally as a device which has the following operational characteristics:

(1) It presents problem materials (questions) to the student;

(2) The student is required to respond to the materials by some overt behavior;

(3) The machine provides the student with knowledge of results of his behavior, usually immediately following each response. This "knowledge of results" consists of telling the student whether his response was correct, or of providing the student with information from which he can evaluate his own response.

Within the generic definition given above there are many dimensions along which one teaching machine may differ from another. The purpose of the present study is to investigate three of these dimensions:

(1) Student response mode. The relative advantages of the multiple choice and the constructed response* modes of operation have been debated at length in the literature on teaching machines. S. L. Pressey (3) has been a leading proponent of the multiple choice mode, while B. F. Skinner (4) has gathered many adherents to the position that constructed response items are superior. Response mode was included in the present study both because of its interest as a variable per se and because of the possibility that it would interact significantly with the other two variables.

(2) Size of item step. Most investigators in the field of automated teaching agree that optimal learning occurs when items presented to the student give the student a large degree of support (1,2,4). According to these investigators the items should be worded and sequenced in such a way that the student finds it simple to proceed from one item to the next, and thus receives a large percentage of positive reinforcements. In the present study, an attempt was made to evaluate experimentally the extent of the purported superiority of small step condition. Certain items were eliminated from one item set to create another set containing fewer items.

---

* The term "constructed response" refers here to the procedure in which the student must complete statements by filling in (constructing) one or more missing words.

Eliminated items covered essentially the same subject material as other items in the set and introduced no new concepts. The derived set may be operationally defined by this process as having larger step size than the original set.

(3) Item sequence control. The great majority of existing teaching machines present subject material items in an essentially predetermined sequence. Relatively little speculation, and even less actual experimentation, have been devoted to the possible advantages of a sequence determined by the behavior of the student.

It is true that most teaching machines in use today have some limited flexibility in the sense that a student must answer a question correctly, or be shown the correct answer, before the next question is presented. Being shown the correct answer or selecting the correct answer after several tries does not insure, however, that the student has actually acquired the necessary knowledge. If he really does understand the concept involved, it is probably inefficient to present many additional questions covering the same concept. If the student does not understand the concept, on the other hand, he should not be presented with more complex questions based upon the original concept, since he is not properly prepared to answer such questions.

A truly adaptable teaching machine should be capable of discriminating critical aspects of the student's response behavior, and selecting subsequent instructional items on the basis of this behavior. Other variables which a teaching machine might modify in response to student behavior include the amount and type of knowledge of results presented to the student, and the number of "cues" given to help the student answer the problem materials. Such a teaching machine would act much more like a human tutor than do most machines available today.

In the present study a simple type of sequence flexibility, which will be called "forward branching," or merely "branching," was compared with the conventional predetermined sequence. The branching procedure (described in more detail in the Method section) was intended to allow a student who has already learned a concept to skip over certain other items covering this same concept. It was hypothesized that students trained with the branching procedure would learn as well as students trained without branching, and would learn in a shorter time due to the omission of some of the items.

METHOD

APPARATUS AND PROCEDURES

The experiment consisted of three phases: (1) A training session for each of 80 subjects with a "teaching machine" operating under one of the eight different teaching procedures being compared; (2) A criterion quiz given to each subject immediately following his training session, to determine how much the subject learned in the training session; and (3) The same criterion quiz, given approximately two weeks after the training sessions to all students in the school classes from which the experimental subjects were drawn. This last phase was intended to provide a measure of retention in the subjects, as well as a comparison of subjects with other students who have not had the experimental training session.

Although the training session for any single subject required less than two hours, the sessions for all subjects extended over a period of one week. During this week the training sessions replaced regular classroom studies for the experimental subjects, while the control group continued to receive its normal instruction. Subject material for the classroom instruction during this week, and up to the final administration of the criterion test, was not closely related to the material taught with the teaching machines.

Six identical sets of equipment were used in the experimental training sessions. The sets actually constituted simulated teaching machines, since for purposes of convenience human experimenters were used in place of automatic control mechanism. Each set consisted of a wooden screen, a number of pushbuttons and lights used for communication between experimenter and subject, a deck of 5" x 8" cards containing instructional items, and four sheets ("panels") with information relating to the instructional items. The wooden screen had a window in its center in which the experimenter placed cards so the instructional items could be seen by the subject. A holder on the experimenter's side of the window supported the cards, and also prevented the subject from observing the experimenter through the window at any time. The information sheets were attached to the subject's side of the screen for all experimental training sessions, and could be used by the subject as an aid in answering the instructional items; subjects could not use the sheets while taking the criterion quiz, however.

Each of the six sets of experimental equipment could be operated under any one of eight teaching procedures. These eight teaching procedures represent the eight combinations of three experimental variables, each variable having two possible values as follows: (1) Response mode (multiple choice versus constructed response); (2) Size of steps between successive items to be taught (small steps versus large steps); and (3) Type of item sequence control (branching versus non-branching). Thus each subject was run under one of the following conditions:

a. Multiple Choice; Small Steps; No Branching
b. Constructed Response; Small Steps; No Branching
c. Multiple Choice; Large Steps; No Branching
d. Constructed Response; Large Steps; No Branching
e. Multiple Choice; Small Steps; Branching
f. Constructed Response; Small Steps; Branching
g. Multiple Choice; Large Steps; Branching
h. Constructed Response; Large Steps; Branching

Subjects and experimenters were randomly assigned to the eight experimental conditions, within the restriction that ten subjects were trained under each condition.

Printed instructions (See Appendix I, II) were read aloud to each subject before the start of the training session. These instructions described the task to be performed by the subject, and informed him that both accuracy and speed of learning were important.

CONSTRUCTED RESPONSE VS. MULTIPLE CHOICE OPERATION. In the constructed response mode of operation there are two pushbuttons on the subject's side of the screen and two small lights on the experimenter's side. The experimenter places in the window of the screen a card containing a constructed response item. This item consists of a statement with one or more blank spaces representing words to be filled in by the subject. The subject writes his answer on a slip of paper and then taps lightly on the screen to indicate that he has made his response. Upon hearing the tap the experimenter removes from the window a cardboard mask, revealing to the subject the correct answer to the item. The subject compares his own answer against the correct answer, and uses one of his two pushbuttons, marked "right" and "wrong" respectively, to activate the corresponding light on the experimenter's side. This indicates to the experimenter whether or not the subject got the item correct.* * The experimenter removes this card and goes on to the next card, continuing until he has finished the deck. He then repeats the entire process, but this time uses only the items missed by the subject on the first trial. The procedure is continued through successive trials until the subject has answered every item correctly.

In the multiple choice mode there are five pushbuttons on the subject's side of the screen, activating five lights on the experimenter's side. These buttons are used by the subject to indicate his answer to the multiple choice items presented to him. In addition there are two pushbuttons used by the experimenter to activate lights on the subject's side, telling the subject whether he has selected a correct or incorrect response. Multiple choice items consist of statements with blank spaces representing missing words. Below the statement are listed two to five possible choices, one of which is to be selected by the subject to answer the item. This selection is accomplished by the subject's pressing the pushbutton having the same letter label as the answer which he wishes to indicate. The various choices are presented to the subject simultaneously with the item, there being no mask of the sort used with constructed response items. The subject must continue to respond to an item until he has selected the correct answer, upon which the experimenter replaces that item with the next in the deck. After he has completed the first trial (e.g., the first run through the deck), the experimenter continues with additional trials as required, each time using only the cards that the subject missed in his first response on the preceding trial.

"SMALL STEPS" VS. "LARGE STEPS." Some subjects were trained with instructional decks containing many (104) items. These decks are called "small step" decks, because they supposedly require little effort on the part of the subject to answer a particular item once he has gone through the preceding items. Other subjects were trained with "large step" decks containing fewer (56) items, and ostensibly requiring more effort to answer successive items.

* An informal check was made to determine the dependability of the subjects' evaluation of their own responses. Several randomly selected subjects were observed during training, and after training their slips were compared with the correct answers. Results of the check indicated that subjects were generally accurate in their evaluations.

From the standpoint of the operation of the teaching machines, there are no procedural differences between "small step" decks and "large step" decks, except in the number of items presented to the subject.

"BRANCHING" VS. "NON-BRANCHING." This variable determines the sequence in which items are presented to the subject. In the "non-branching" mode, the experimenter simply goes through all the cards of the deck in order. In the "branching" mode, certain items are removed from the deck if the subject answers certain other items correctly on the first try. The items to be removed are taken out of the deck by a second experimenter, so they are never presented to the subject. Thus, in the "branching" mode, the exact number and sequence of items are not fixed, but depend on the performance of the subject. A more complete description of the criteria for omitting items is given in the following section on Subject Material.

PERFORMANCE MEASURES. The following performance measures were recorded for each subject, regardless of mode of instruction:

(1) Time required to finish training.
(2) Score on multiple choice portion of criterion test.
(3) Score on constructed response portion of criterion test.

SUBJECT MATERIAL

The eight item sets (decks) and the three instructional sheets used in the experiment were based on a portion of a college course in elementary psychology used at Harvard University. The first item set obtained was the constructed response, small step deck, without branching. This set consisted of the first 104 items from a larger series of items developed at the Harvard University Psychological Laboratory. Samples of the items are shown in Appendix VI. The remaining seven sets represent combinations of the following transformations of the original items:

(1) A number of possible alternative answers were provided for each item, to make the multiple choice decks.

(2) Certain items were eliminated from the original set of 104 items, to make the "large step" decks of 56 items. The items eliminated were judged to be largely redundant, in the sense that they related to concepts already covered by other items, and added little extra information.

(3) To make the "branching" decks, coded instructions for the branching procedure were marked on the backs of certain cards. These instructions indicated which cards were to be eliminated from the deck during the experiment if the subject correctly answered the marked cards. For convenience, a marked card will be referred to as a "branch card"; the cards to be eliminated when a "branch card" is correctly answered will be called "conditional skip" cards. In general, a "branch card" was the first of several items in the deck covering a particular concept, while the associated "conditional skip" cards consisted of the remaining items covering essentially the same concept. "Small step" branching decks included 13 "branch cards" and 39 "conditional skip" cards, while "large step" branching decks included 12 "branch cards" and 22 "conditional skip" cards.

The three instructional sheets, or "panels," used in the study covered most of the major concepts presented in the instructional items, but in the form of definitions and brief descriptive materials. Appendix V contains several examples of statements from the panels, which were prepared by Harvard University.

CRITERION TEST

The criterion test consisted of 36 questions, of which 19 were constructed response (fill-in) and 17 were multiple choice. All of the questions were based upon the material contained in the original (constructed response; small step; without branching) set of instructional items. The person who prepared the criterion items had not seen the other seven item sets prepared for instruction; conversely, the persons preparing the alternate item sets had not seen the criterion questions.

None of the criterion questions were duplicates of instructional items, though many of the same words were used as correct responses. Most of the criterion questions were of the application type, in which a situation is presented and the respondent identifies the principle involved or attempts to explain or predict an outcome. Examples of criterion test questions are shown in Appendix IV.

Reliability of the criterion test, as estimated by the application of Kuder-Richardson Formula 20 to the scores of the experimental subjects, was as follows: whole test, .89; nstructed response portion, .85; multiple-choice portion, .79. Test-retest correlations for experimental subjects were .81 for the whole test, .79 for the constructed response portion, and .58 for the multiple choice portion.

EXPERIMENTAL SUBJECTS

The 80 experimental subjects and the 104 members of the control group were taken from beginning psychology classes in Santa Monica City College. They had been in the psychology course for about one month, using F. Ruch's Psychology and Life as a text and covering some general background in the area of psychology, but had not been given specific instruction in any material closely related to the subject material used in the experiment.

RESULTS

1. A psychology pre-test (See Appendix III) was given to all experimental subjects and also to the students who had not been selected for the experiment. Means and SD's for the two groups are given in Table I. As shown in Table II, no significant difference was obtained between the 80 experimental S's and the 104 non-experimental S's using a "t" test. The highest possible score on the pre-test was 61.

After the experimental training period a criterion test, covering applications of the concepts that were being taught, was administered to experimental and non-experimental groups. The criterion test consisted of two subtests, one with multiple choice items, the other with constructed response (fill in) items.

Table I gives the mean scores and SD's for the experimental and control groups. Maximum possible scores on the criterion were 19 and 17 for the constructed response and multiple choice subtests respectively.

The experimental group was superior to the control group on the total criterion and on both the multiple choice and constructed response subtests, as shown in Table II. Differences were significant at the .01 level.

3. The criterion test was re-administered to the experimental S's three weeks after the first administration as a measure of retention. Means and SD's for the experimental group retest are given in Table I. As shown in Table II, no significant difference was obtained on the total criterion (multiple choice plus constructed response), when the mean of the first administration was compared with the mean of the second administration. On the multiple choice criterion subtest, however, retest scores were significantly higher (.01 level) than original test scores. This improvement in performance may be attributable to the fact that students became acquainted with the criterion questions, although it is difficult to explain why a similar improvement did not occur on the constructed response subtest.

The evidence from these first three analyses indicates that the automated teaching materials used with the experimental subjects resulted in significant learning of the concepts taught and that this learning is retained at least for a three week period. These findings do not mean that the experimental S's exceeded groups receiving conventional classroom instruction on the same materials, since the control S's were not being taught the same concepts as the experimental group. The purpose of the experimental versus control group comparison was to determine whether <u>any</u> learning was taking place in the experimental group over that which might have occurred without such training.

TABLE I

MEANS AND STANDARD DEVIATIONS FOR EXPERIMENTAL AND CONTROL GROUPS ON PRE-TEST AND CRITERION SUBTESTS

SOURCE OF SCORES		Experimental Group		Control Group	
		Mean	SD	Mean	SD
PRE-TEST		45.5	8.0	44.4	7.7
FIRST ADMINISTRATION OF CRITERION TEST	Multiple Choice	11.6	2.5	10.5	2.6
	Constructed Response	13.5	3.9	8.6	3.6
	Total Criterion	25.1	5.9	19.1	5.2
SECOND ADMINISTRATION OF CRITERION TEST	Multiple Choice	12.4	2.5	NOT ADMINISTERED	
	Constructed Response	13.4	3.4		
	Total Criterion	25.8	5.3		

TABLE II

COMPARISONS FOR EXPERIMENTAL AND CONTROL GROUPS ON PRE-TEST AND CRITERION TEST

COMPARISON		t	df	P
EXPERIMENTAL GROUP MINUS CONTROL GROUP ON PRE-TEST		1.00	182	not signif.
EXPERIMENTAL GROUP MINUS CONTROL GROUP ON ORIGINAL CRITERION TEST	Multiple Choice Subtest	2.99	195	.01
	Constructed Response Subtest	9.07	195	.01
	Total Criterion	7.35	195	.01
ORIGINAL CRITERION TEST MINUS RETEST FOR EXPERIMENTAL GROUP	Multiple Choice Subtest	2.92	70	.01
	Constructed Response Subtest	.64	70	not signif.
	Total Criterion	1.67	70	not signif.

4. Table III contains pre-test scores and criterion test scores for the eight experimental treatment groups. The correlations between pre-test scores and criterion scores were as follows:

Correlation between pre-test and constructed response scores = .411 (.01 level of significance)

Correlation between pre-test and multiple choice scores = .404 (.01 level)

Correlation between pre-test and total criterion = .442 (.01 level)

5. Results of covariance analysis of the constructed response criterion measures for the eight experimental groups are shown in Table IV. Pre-test scores were used as a control variable in this analysis because of the significance of the correlation between pre-test and constructed response criterion scores. A significant main effect was found for the size of step factor (.05 level of significance), in the direction of higher scores for small step trainees. A significant interaction was also found between the mode of response and branching variables (.05 level). The assumption of homogeneity of regression was tested and found to be satisfied.

Table V gives observed mean constructed response scores for comparison of the effects of the three experimental variables (response mode, step size and branching procedure). Mean constructed response criterion scores are also given for the four combinations of response mode and branching procedure, since this interaction was found significant in the covariance analysis.

The right-hand column of Table V shows the constructed response criterion means adjusted for pre-test scores.

6. Covariance analysis of the multiple choice criterion measures of the subjects in the eight experimental groups was also performed with psychology pre-test scores as a control variable. The results of this analysis, summarized in Table VI, yielded no significant main effects or interactions among the experimental groups. Table VII shows the observed and adjusted multiple choice criterion scores for the main treatments.

7. Covariance analysis of the total criterion measures (multiple choice plus constructed response), using the psychology pre-test as a control variable, is summarized in Table VIII. No significant main effects or interactions were obtained among the experimental groups. Observed and adjusted mean scores for the experimental groups on the total criterion test are shown in Table IX.

8. An analysis of variance of differences among the eight experimental groups with respect to the amount of time taken to complete training was also performed. All three main effects are significant at the .01 level. The constructed response training condition took more time than the multiple choice condition; the small step condition took more time than the large step condition; and the non-branching condition took more time than the branching condition. Interactions were not significant. The mean scores on the time criterion are shown in Table X and the analysis of variance is summarized in Table XI.

TABLE III

MEAN SCORES FOR EIGHT EXPERIMENTAL GROUPS ON PRE-TEST AND ON MULTIPLE CHOICE AND CONSTRUCTED RESPONSE CRITERIA

TREATMENT GROUPS			MEAN ON PRE-TEST	OBSERVED MEAN CRITERION SCORES		
Response Mode	Step Size	Branching		MC	CR	TOTAL
1. CR	Small	Branch	49.4	12.1	13.8	25.9
2. MC	Small	Branch	45.0	12.1	14.3	26.4
3. CR	Large	Branch	45.0	11.9	13.3	25.2
4. MC	Large	Branch	47.4	10.6	13.2	23.8
5. CR	Small	No Branch	41.2	11.6	15.1	26.7
6. MC	Small	No Branch	45.3	11.9	13.6	25.6
7. CR	Large	No Branch	45.5	12.2	14.0	26.2
8. MC	Large	No Branch	45.3	10.6	10.7	21.3
ALL GROUPS			45.5	11.6	13.5	25.1

TABLE IV

COVARIANCE ANALYSIS OF CONSTRUCTED RESPONSE CRITERION SCORES USING A PSYCHOLOGY PRE-TEST AS A CONTROL

SOURCE OF VARIATION	ADJUSTED SUM OF SQUARES	df	ms	F
A. Mode of Response	29.22	1	29.22	2.48
B. Size of Step	46.99	1	46.99	3.99 *
C. Branching	1.27	1	1.27	.11
AB	11.05	1	11.05	.94
AC	53.28	1	53.28	4.52 *
BC	18.34	1	18.34	1.56
ABC	2.23	1	2.23	.19
Within Cells	846.88	71	11.93	
TOTAL	1009.26	78		

* .05 Level

TABLE V

OBSERVED AND ADJUSTED MEAN SCORES ON CONSTRUCTED RESPONSE CRITERION SUBTEST

GROUPS (Training Mode)	OBSERVED MEANS ON CONSTRUCTED RESPONSE CRITERION	CONSTRUCTED RESPONSE MEANS ADJUSTED FOR PRE-TEST SCORES
Constructed Response	14.05	14.14
Multiple Choice	12.95	12.86
Small Step	14.20	14.31
Large Step	12.80	12.69
Branching	13.65	13.21
No-Branching	13.35	13.79
Branching with Constructed Response	13.55	12.93
Branching with Multiple Choice	13.75	13.50
No-Branching with Constructed Response	14.55	15.35
No-Branching with Multiple Choice	12.15	12.92

TABLE VI

COVARIANCE ANALYSIS OF MULTIPLE CHOICE CRITERION SCORES USING A PSYCHOLOGY PRE-TEST AS A CONTROL

SOURCE OF VARIATION	ADJUSTED SUM OF SQUARES	df	ms	F
A. Mode of Response	10.22	1	10.22	1.94
B. Size of Step	9.20	1	9.20	1.74
C. Branching	.99	1	.99	.19
AB	15.66	1	15.66	2.97
AC	.82	1	.82	.16
BC	.03	1	.03	.01
ABC	1.03	1	1.03	.20
Within Cells	374.81	71	5.28	
TOTAL	412.76	78		

T A B L E VII

OBSERVED AND ADJUSTED MEAN SCORES ON MULTIPLE CHOICE CRITERION SUBTEST

GROUPS (Training Mode)	OBSERVED MEANS ON MULTIPLE CHOICE CRITERION	MULTIPLE CHOICE MEANS ADJUSTED FOR PRE-TEST SCORES
Constructed Response	11.95	11.97
Multiple Choice	11.30	11.34
Small Step	11.93	11.97
Large Step	11.33	11.89
Branching	11.68	11.66
No Branching	11.58	11.60

TABLE VIII

COVARIANCE ANALYSIS OF TOTAL CRITERION SCORES USING A PSYCHOLOGY PRE-TEST AS A CONTROL VARIABLE

SOURCE OF VARIATION	ADJUSTED SUM OF SQUARES	df	ms	F
A. Mode of Response	72.15	1	72.15	2.65
B. Size of Step	99.96	1	99.96	3.68
C. Branching	4.65	1	4.65	.17
AB	54.73	1	54.73	2.01
AC	65.84	1	65.84	2.42
BC	18.35	1	18.35	.67
ABC	5.57	1	5.57	.20
Within Cells	1930.80	71	27.19	
TOTAL	2252.05	78		

TABLE IX

OBSERVED AND ADJUSTED MEAN SCORES ON TOTAL CRITERION TEST

GROUPS (Training Mode)	OBSERVED MEANS ON TOTAL CRITERION	TOTAL CRITERION MEANS ADJUSTED FOR PRE-TEST SCORES
Constructed Response	26.00	26.02
Multiple Choice	24.28	24.24
Small Step	26.15	26.19
Large Step	24.13	24.09
Branching	25.32	25.30
No Branching	24.95	24.97

T A B L E X

MEAN SCORES ON TRAINING TIME CRITERION

GROUPS	MINUTES
Constructed Response	54.4
Multiple Choice	44.4
Small Step	57.3
Large Step	41.5
Branching	43.8
No Branching	55.0

T A B L E XI

ANALYSIS OF VARIANCE OF TIME REQUIRED TO COMPLETE TRAINING

SOURCE OF VARIATION	ss	df	ms	F
A (Mode of Response)	2000.00	1	2000.00	11.66 **
B (Size of Step)	4992.80	1	4992.80	29.11 **
C (Branching)	2508.80	1	2508.80	14.62 **
AB	42.05	1	42.05	.25
AC	48.05	1	48.05	.28
ABC	540.80	1	540.80	3.15
Within Cells	12347.80	72	171.50	
TOTAL	22642.75	79		

* * .01 Level

DISCUSSION

CONSTRUCTED RESPONSE CRITERION DATA

SIZE OF STEP VARIABLE: The finding that treatment groups given more items (small steps) learned more than groups receiving few items is in accord with B. F. Skinner's emphasis on the importance of small steps in writing instructional item sets. The superiority of small steps in total amount learned must be weighed, however, against the significantly greater training time required by small step trainees than by large step trainees. The best choice of item step size for any applied teaching situation will be determined in part by the extent to which required training time is a critical consideration.

BRANCHING VARIABLE: In the comparison of large item steps and small steps it was found that small step trainees, who were given more items, learned more than large step trainees. In view of this finding it is notable that branching, with its consequent reduction in total number of items presented, did not differ significantly from non-branching in amount taught. One possible explanation lies in the fact that the difference in the number of items given to branching and non-branching trainees was less than the difference between large and small step trainees. A more important factor, however, may have been the way in which items were eliminated in the branching procedure. An item was skipped only after the trainee had demonstrated some knowledge of the concept taught by that item, with the result that every major concept was covered by at least one item. At the same time, the branching procedure made possible a significant decrease in required training time, as compared with the non-branching procedure.

When both the amount learned and the required training time are considered, the branching procedure appears to offer an over-all advantage over non-branching. The present investigators feel, moreover, that branching, or other types of control flexibility, offer even greater potential for future improvement teaching machine effectiveness. In the study reported here, limitations due to human simulation of the teaching machine control mechanism placed heavy restrictions on the branching procedure. The greatly simplified skipping procedure resulted in a significant advantage in training time but not in amount learned. More complex forms of control flexibility, using some type of automatic equipment, may prove superior to pre-determined sequence control in both amount learned and speed of learning.

RESPONSE MODE VARIABLE: Under the conditions of the experiment reported here, student response mode did not significantly affect the amount learned by the students. Since required training time was significantly less for multiple choice trainees than for constructed response trainees, the over-all advantage appears to be with the multiple choice mode. It is possible, of course, that the results would be different under other experimental conditions, and further research in the area may prove valuable.

An important consideration in the evaluation of response mode for applied purposes is the ease with which a particular mode can be incorporated into an actual teaching machine. This practical consideration must be weighed along

with the teaching effectiveness of the response mode. The use of the multiple choice mode, rather than the constructed response mode, greatly simplifies the design of any teaching device which is intended to evaluate student responses automatically.

RESPONSE MODE-BRANCHING INTERACTION: The significant interaction between response mode and branching procedure can probably be attributed to the combined effects of two factors, neither of which is sufficiently powerful to have statistically significant effect when taken separately. The first factor contributing to the interaction is the number of items presented to the subjects. By this hypothesis the constructed response trainees learned more without branching than with branching because the branching procedure caused a reduction in the total number of items. The second factor contributing to the interaction effect was the relation between the response mode required in the training session, and that required in the criterion test. As previously noted, all the criterion scores being compared here were obtained from the constructed response portion of the criterion test. This might well be expected to give an advantage to the constructed response trainees due to generalization decrement on the part of the multiple choice trainees.

Table V shows that main effect differences in favor of the non-branching and constructed response conditions, though non-significant, tend by their direction to support the "combined effects" interpretation presented here.

MULTIPLE CHOICE CRITERION DATA

Although the mean of the experimental S's on the multiple choice criterion significantly exceeded the control group mean, no significant differences were obtained among the different experimental groups. Differential effects of the various experimental treatment combinations on the multiple choice criterion appear to have been masked by a larger effect common to all experimental groups. This masking error variance can probably be attributed to the nature of the multiple choice criterion subtest. Of the 17 multiple choice items, nine had four alternatives, two had three alternatives, and six had only two alternatives. The probability of obtaining correct answers on a chance basis could account on the average for approximately one third of the items. The largest portion of the variance of the scores on the multiple choice test may be due to this chance factor. Students of the sort used in the study are usually highly skilled in picking out subtle specific determiners by which they can eliminate implausible alternatives on grounds other than factual content. These skills, which together with day-to-day fluctuations in individual performance are treated as systematic variance in the Kuder-Richardson method of estimating reliability, contribute to over-estimation of the effective reliability of the test. In this case, the Kuder-Richardson reliability estimate was .79. A more realistic estimate of reliability is given by the multiple choice criterion test-retest correlation, which was only .58.

An alternative to the "masking" interpretation is that there was actually no difference in effectiveness among the experimental treatments. The results on the constructed response criterion and the operational differences among the treatments, however, tend to place doubt on this second hypothesis. Although not significant, the direction of differences among the means of the experimental groups shown in Table VII tend to support the results obtained on the constructed response criterion.

TOTAL CRITERION TEST DATA

The results of the analysis of the total criterion follow from the results obtained on the analyses of the two subtests. Error variance in the multiple choice portion of the total criterion appears to have masked inter-group differences so that these differences are not significant. It can be noted in Table VIII that the covariance F test approaches but does not reach significance at the .05 level for both the size of the step factor and the mode of response - branching interaction. The direction of the means in Table IX also agrees with differences obtained on the analysis of scores on the constructed response criterion.

TIME CRITERION

Differences in required training time are generally in the expected direction. Small step groups and non-branching groups took longer than large step and branching groups because they were required to answer more questions. The fact that constructed response groups took longer than multiple choice groups indicates that the composition and writing of answers is more time consuming than the recognition of a "correct" solution among several alternatives, even though multiple choice trainees were sometimes required to make several selections before choosing the right answer.

SUMMARY

Eight groups of ten junior college students were given an experimental training session with manually simulated teaching machines, each group being taught with a different mode of teaching machine operation. The three independent variables were student response mode, size of steps between successive items and sequencing (branching) procedure. A written criterion test was given to all subjects immediately after the training session, and again three weeks later. The same criterion test was given to a control group which had no training with the teaching machine, but which came from the same school classes as the experimental subjects. Dependent variables were the required teaching machine training time and scores on the criterion test. Scores on a pre-test were used as a control variable in a covariance design. Major results were as follows:

(1) Use of the simulated teaching machines led to significant learning by the subjects, as determined by comparison with the control group.

(2) The multiple choice response mode took significantly less time than the constructed response mode. No significant difference was obtained between response modes on the criterion test.

(3) Small item steps required significantly more training time, but also yielded significantly higher test scores than large item steps on the constructed response criterion subtest.

(4) The branching conditions required less training time than non-branching, but were not significantly different on the criterion test. A significant interaction was obtained between the mode of response and branching variables on the constructed response criterion. This interaction resulted from a high mean criterion score obtained by the constructed response - non-branching group.

(5) No significant differences were obtained among the experimental groups on the multiple choice criterion subtest, or on the total (multiple choice plus constructed response) criterion test.

REFERENCES

1. Glaser, R. & Homme, L. E., Relationships between the programmed textbook and teaching machines. Paper read at Air Force Office of Scientific Research and University of Pennsylvania Conference on The Automatic Teaching of Verbal and Symbolic Skills, December, 1958.

2. Porter, D., Teaching Machines. Harvard Graduate Sch. Educ. Bull., 1958, 3, (1).

3. Pressey, S. L., Development and appraisal of devices providing immediate automatic scoring of objective tests and concomitant self-instruction. J. Psychol., 1959, 29, 417-447.

4. Skinner, B. F., Teaching Machines. Science, 1958, 128, 969-977.

APPENDICES

APPENDIX I	Directions for Subjects - Constructed Response Items
APPENDIX II	Directions for Subjects - Multiple Choice Items
APPENDIX III	Pre-Test Questions (Examples)
APPENDIX IV	Criterion Test Questions (Examples)
APPENDIX V	Instruction Panels (Examples)
APPENDIX VI	Training Items (Original Constructed Response, Small Step Form): Examples

APPENDIX I

DIRECTIONS FOR SUBJECT - CONSTRUCTED RESPONSE ITEMS

The object of this experiment is not to find out how much you know, or to compare you with any other student, but to find out what teaching procedure does the best job of teaching you something about psychology.

You will be shown a series of questions. Each question will contain a statement with one or more blank spaces. You are to write your answers on a sheet of blank paper pad in front of you. As soon as you have written your answers you tap the screen with your pencil, indicating that you are ready to see the correct answer. The correct answer will then be revealed in the window beneath the original question. If your answer was correct, you press the pushbutton marked "right." If your answer was wrong, you press the "wrong" button. Hold the "right" or "wrong" button down until the next card appears. Then tear off the sheet with your answer on it, and drop it in the box in front of you. You are then ready to go through the answering procedure with the next card just as you did with the first card.

It is up to you to decide whether you have answered a particular question correctly. If a card has more than one blank, you must answer all parts correctly or the whole question must be counted as "wrong." In order to be marked "right," your answer must use exactly the same word or words as the correct answer, or words that mean the same thing.

A record will be kept of how much you learn, and of the time required for you to learn it. It will be up to you to decide how much time you should spend in trying to answer a particular question correctly.

APPENDIX II

DIRECTIONS FOR SUBJECT - MULTIPLE CHOICE ITEMS

The object of this experiment is not to find out how much you know or to compare you with any other students, but to find out what teaching procedure does the best job of teaching you something about psychology.

You will be shown a series of multiple choice questions. Read the question, then indicate your answer by pressing the pushbutton corresponding to your choice. For example, if the answer you choose is alternative B, press pushbutton B until you see either a red light or a green light. The green light means that you were correct. The red light means that you were wrong and you must try again until you get it right. This same procedure will be used for the next card and the next, until you have finished.

A record will be kept of how much you learn, and of the time required for you to learn it. It will be up to you to decide how much time you should spend in trying to answer a particular question correctly.

APPENDIX III

EXAMPLES OF PRE-TEST QUESTIONS

1. The humoral theory was proposed by: (a) Kretschmer;(b) Mesmer; (c) Hippocrates; (d) Sheldon.

2. The oldest known characterological theory is: (a) phrenology; (b) humoral theory; (c) somatotype theory; (d) psychophysics.

3. Sheldon is chiefly known for his connection with: (a) the somatotype theory; (b) the humoral theory; (c) phrenology; (d) psychophysics.

4. A person who believed in phrenology would attempt to learn about personalty traits by studying: (a) body fluids; (b) body build; (c) reactions to specific stimuli; (d) the shape of the skull.

5. Which of the following would be most likely to fluctuate between moods of happiness and depression: (a) schizothymic type; (b) leptosome; (c) Mesomorph; (d) cyclothymic type.

6. Which of the following would probably be most interested in athletics and other muscular activities: (a) endomorph; (b) mesomorph; (c) leptosome; (d) pyknic type.

7. An imbalance of magnetic fluids in the body was regarded as the cause of illness by: (a) Mesmer; (b) Hippocrates; (c) Kretschmer; (d) Fechner.

8. Mesmerism led to the development of the technique known as: (a) projective testing; (b) shock therapy; (c) hypnosis; (d) the clinical method.

9. Philippe Pinel is noted primarily for his work in: (a) experimental physiology; (b) the mind-body problem; (c) psychophysics; (d) the reform of mental hospitals.

10. The father of psychophysics is: (a) Mesmer; (b) Pinal; (c) Helmholtz; (d) Fechner.

APPENDIX IV

EXAMPLES OF CRITERION QUESTIONS

Directions: On the multiple choice items, put a check (X) on the line to the left of the correct answer. On the incomplete sentences, print the correct answer in the space provided.

1. Bill, a candidate for the draft, was tapped on the knee with a rubber hammer; his leg jerked. After waiting five minutes the doctor tapped his knee again, but harder. The movement of his leg the second time should be:

_____a) greater
_____b) the same
_____c) smaller

2. Which statement is correct?

_____a) The word "reflex" means the same as"response."
_____b) Any reflex action is a response to a stimulus.
_____c) Any response to a stimulus is a reflex action.
_____d) None of the above statements is correct.

3. Which of the following is a stimulus?

_____a) Saliva in your mouth during a meal.
_____b) The sound of a pencil hitting the floor.
_____c) Picking up a pencil that has fallen.
_____d) A sound that is not heard.

4. When a loud noise occurs near an infant he will normally jerk his body. This combination is called the "startle_____."

5. A reflex response always comes_____the stimulus.

6. When a conditioned stimulus precedes a response we say that the stimulus_____the response.

APPENDIX V

EXAMPLES OF INSTRUCTIONAL PANELS

A light (a stimulus) flashed into the eye causes the pupil to contract (a response). This relation between the stimulus and its response is called the pupillary reflex. The time elapsing between the onset of the stimulus and the onset of the response is the latency of the reflex. The lowest intensity of light which will elicit contraction is called the threshold of the stimulus.

NEW TECHNICAL WORDS

stimulus

response

reflex

latency

elicit

threshold

APPENDIX VI

EXAMPLES OF TRAINING ITEMS

1. A doctor taps your knee (patellar tendon) with a rubber hammer to test your "_____." "reflexes"

2. If your reflexes are normal you _____ to the tap on the knee with a knee jerk. "respond" (or react)

3. In the knee jerk reflex, the knee jerk is the _____. response (or reaction)

4. The stimulus which elicits the knee jerk response is the _____. tap (on the knee)

5. To explain the knee jerk you need only describe the _____. stimulus (or tap on the knee)

6. The tap on the knee _____ the knee jerk response. elicits (or causes, produces, brings about)

7. In the knee jerk reflex the temporal order is first the (1) _____ then the (2) _____.
   (1) stimulus (or tap)
   (2) response (or knee jerk)

8. The interval of time which necessarily elapses between the stimulus and response is called the latency. The time between tapping the knee and the knee jerk is the _____ of the knee jerk reflex. latency

9. The minimum intensity of a stimulus just sufficient to elicit a response is called the threshold of the stimulus. A tap on the knee will elicit the knee jerk only if the tap exceeds the _____. threshold

10. A very hard tap on the knee elicits a(n) (1) _____ knee jerk while a tap barely above the threshold elicits a(n) (2) _____.
    (1) large (or big, intense)
    (2) small (or little, feeble)